SAS/STAT® 13.1 User's Guide
What's New in SAS/STAT 13.1

Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

Chapter 1
What's New in SAS/STAT 13.1

Contents

Overview

SAS/STAT 13.1 includes several new procedures and many enhancements.

New Procedures

Experimental BCHOICE Procedure

The experimental BCHOICE procedure performs Bayesian analysis for discrete choice models. Discrete choice models are used in marketing research to model decision makers' choices among alternative products and services. The decision makers might be people, households, companies and so on, and the alternatives might be products, services, actions, or any other options or items about which choices must be made (Train 2009). The collection of alternatives that are available to the decision makers is called a choice set. Discrete choice models are derived under the assumption of utility-maximizing behavior by decision makers. When individuals are asked to choose among a set of alternatives, they usually determine the level of utility that each alternative offers.

To use the BCHOICE procedure, you need to specify the model for the data. You can also supply a prior distribution for the parameters if you want something other than the default noninformative prior. PROC BCHOICE obtains samples from the corresponding posterior distributions, produces summary and diagnostic statistics, and saves the posterior samples in an output data set that can be used for further analysis.

ICLIFETEST Procedure

The ICLIFETEST procedure performs nonparametric survival analysis for interval-censored data. You can use the ICLIFETEST procedure to compute nonparametric estimates of survival functions and to examine the equality of survival functions via statistical tests. The ICLIFETEST procedure is similar to the LIFETEST procedure. The two procedures share the same analytic objectives: estimating and summarizing subjects' survival experiences and comparing them systematically. The distinction between these procedures lies in the types of data that they are designed to handle. The ICLIFETEST procedure is intended primarily for handling interval-censored data, whereas the LIFETEST procedure deals exclusively with right-censored data. You can use the ICLIFETEST procedure to analyze data that are left-censored, interval-censored, or right-censored. However, if the data to be analyzed contain only exact or right-censored observations, you should use the LIFETEST procedure because it provides specialized methods for right-censored data.

Experimental IRT Procedure

The experimental IRT procedure fits item response models. These models are widely used in education to calibrate and evaluate items in tests, questionnaires, and other instruments and to score subjects on their abilities, attitudes, or other latent traits. In recent years, IRT models have also become increasingly popular

in health behavior, quality of life, and clinical research. The IRT procedure fits the Rasch model; one-, two-, three-, and four-parameter models; and the graded response model with a logistic or probit link. It enables different items to have different response models, performs multidimensional exploratory and confirmatory analysis, performs multiple-group analysis, and estimates factor scores.

Highlights of Enhancements

The following are highlights of other enhancements in SAS/STAT 13.1:

- The MI procedure now provides the MNAR statement to facilitate sensitivity analysis.

- The Tweedie distribution is now supported by the GENMOD procedure.

- The competing risk model of Fine and Gray (1999) is available in the PHREG procedure.

- With the NLIN procedure, you can generate both bootstrap estimates of confidence intervals for the parameters and bootstrap estimates of the covariance matrix and correlation matrix of the parameter estimates.

- The MCMC procedure is now multithreaded.

- Path diagrams are available with the CALIS procedure.

- You can now compute power for PROC GLM–type MANOVA and repeated measurements with the GLMPOWER procedure.

- The SURVEYMEANS procedure produces domain quantile estimates.

More information about the changes and enhancements follows. Details can be found in the documentation for the individual procedures in the *SAS/STAT 13.1 User's Guide*.

Highlights of Enhancements in SAS/STAT 12.1 and SAS/STAT 12.3

Some users might be unfamiliar with updates made in the previous releases. SAS/STAT 12.1 introduced the ADAPTIVEREG, QUANTLIFE, QUANTSELECT, and STDRATE procedures. What follows are highlights of the other enhancements provided with SAS/STAT 12.1:

- The MCMC procedure models missing values by default. The RANDOM statement supports multilevel hierarchy to an arbitrary depth. The procedure also implements faster and more efficient sampling algorithms.

- The PHREG procedure supports Bayesian frailty models.

- The FMM procedure for finite mixture models is production and adds several truncated distributions.

- The LIFEREG and PROBIT procedures include additional postprocessing statements. They now support the TEST, LSMEANS, LSMESTIMATE, ESTIMATE, SLICE, and EFFECTPLOT statements.

- The FREQ procedure produces mosaic plots.

- The SURVEYSELECT procedure provides Poisson sampling.

- The SURVEYMEANS procedure performs poststratification estimation.

- The GLM, MIXED, GLIMMIX, and ORTHOREG procedures support the REF= option in the CLASS statement.

- The CALIS procedure supports robust estimation and produces case-level residual analysis with graphical output.

SAS/STAT 12.3 was primarily a maintenance release that was made available with Base SAS 9.4. However, it did introduce the HPGENSELECT procedure, which provides model selection for generalized linear models. PROC HPGENSELECT is documented in *SAS/STAT User's Guide: High-Performance Procedures*.

Enhancements

ADAPTIVEREG Procedure

The ADAPTIVEREG procedure is production in this release.

CALIS Procedure

The CALIS procedure now includes path diagrams, which provide visually informative representations of the interrelationships among variables in structural equation models. PROC CALIS provides path diagrams for the initial model specification, the unstandardized solution, or the standardized solution.

FMM Procedure

The new criterion panel graph displays the progression of model fit criteria for mixtures that have different numbers of components. In addition, the KMIN= and KMAX= options are now available for Bayesian analyses.

FREQ Procedure

The new OR(CL=SCORE) option provides score confidence limits for the odds ratio. The RISKD-IFF(COMMON) option provides Mantel-Haenszel and summary score estimates of the common risk (proportion) difference. The score confidence limits can be displayed in the odds ratio plot, and the common risk difference can be displayed in the risk difference plot.

The MIDP option in the EXACT statement produces mid *p*-values for exact tests.

The new COLORSTAT= option for mosaic plots colors the tiles according to the values of the Pearson residuals or the standardized residuals. The SCALE=GROUPPERCENT option for two-way frequency plots displays the row or column percentages (instead of the overall percentages). The CLDISPLAY=SERIFARROW and CLDISPLAY=LINEARROW options are now available to control the error bars in odds ratio, relative risk, risk difference, and kappa plots.

The CROSSLIST(PEARSONRES) option displays the Pearson residuals in the CROSSLIST table.

GAM Procedure

The REF= option in the CLASS statement enables you to designate a specific reference level for a classification variable.

GENMOD Procedure

The DIST=TWEEDIE option in the MODEL statement enables you to specify the Tweedie distribution.

GLIMMIX Procedure

Weighted multilevel models are now available. These models are very useful in analyzing survey data that arise from multistage sampling. In these sampling designs, survey weights are often constructed to account for unequal sampling probabilities, nonresponse adjustments, and poststratification.

GLMPOWER Procedure

You can now perform power analyses for multivariate models, including PROC GLM-style MANOVA and repeated measurements analysis.

GLMSELECT Procedure

The SELECTION=ELASTICNET option specifies the elastic net method, an extension of LASSO that estimates parameters based on a version of ordinary least squares in which both the sum of the absolute regression coefficients and the sum of the squared regression coefficients are constrained. If the model contains classification variables, then these corresponding effects can be split.

The CHOOSE=CVEX suboption of the SELECTION option specifies the predicted residual sum of square with *k*-fold external cross validation as the criterion for choosing the model. The STOP=L1 suboption of the SELECTION option is available for SELECT ION=LASSO or SELECTION=ELASTICNET; it stops selection at the step where the L1 criterion is equal to the value specified by the L1=value option.

LIFETEST Procedure

Chapter 23, "Customizing the Kaplan-Meier Survival Plot," provides examples of using procedure options that enable you to modify the Kaplan-Meier survival plot in PROC LIFETEST. This chapter also provides a set of macros that make it easy to perform extensive customizations of the survival p lot.

MCMC Procedure

The MCMC procedure is now multithreaded and can take advantage of multiple processors. The NTHREADS= option in the PROC MCMC statement specifies the number of threads for simulation. When sampling model parameters, PROC MCMC allocates data into different threads and calculates the objective function by accumulating values from each one. When sampling random-effects parameters and missing data variables, PROC MCMC generates a subset of these parameters on individual threads simultaneously at each iteration. Most sampling algorithms are threaded. By default, NTHREADS=1.

PROC MCMC now permits parameters (or functions of parameters) in all truncated distributions (LOWER= and UPPER= options) in both the PRIOR and the MODEL statements.

MI Procedure

The new MNAR statement facilitates sensitivity analysis by generating multiple imputations for different scenarios under the assumption that the data are missing not at random. The MNAR statement is used in conjunction with the MONOTONE and FCS statements .

The options LINK=LOGIT and LINK=GLOGIT are now included in the FCS and MONOTONE statements and provide generalized logistic regression.

MIANALYZE Procedure

The MIANALYZE procedure now reads results from generalized logistic regression.

NLIN Procedure

The BOOTSTRAP statement requests bootstrap estimation of confidence intervals for parameters and bootstrap estimates of the covariance matrix and the correlation matrix of the parameter estimates. PROC NLIN also produces histograms and scatter plots of the bootstrap parameter estimates.

NPAR1WAY Procedure

The MIDP option in the EXACT statement requests exact mid *p*-values for the exact tests.

PHREG Procedure

The PHREG procedure now provides the competing risk model of Fine and Gray (1999) when you specify the EVENTCODE= option in the MODEL statement. The EV option in the PROC PHREG statement requests the Schemper and Henderson (2000) measure of explained variation. The ZPH option requests a test of nonproportional hazards based on the scaled Schoenfeld residuals; in addition, a plot of the time-varying coefficients is displayed with a smooth curve for each predictor.

PLM Procedure

The SCORE statement enables you to score new observations for zero-inflated models.

PROBIT Procedure

The CLASS statement now includes the REF= , ORDER= , and DESCENDING= options. The MODEL statement now includes the EVENT= and DESCENDING= option for the response variable.

QUANTLIFE Procedure

The QUANTLIFE procedure is now production.

QUANTSELECT Procedure

The QUANTSELECT procedure is now production.

REG Procedure

The REG procedure displays fit and residual plots as heat maps rather than as scatter plots when the number of observations is large and ODS Graphics is enabled.

SEQDESIGN Procedure

The CEILING= option specifies the additional sample size information to be displayed in the "Number of Events (D) and Sample Sizes (N)" table. The CEILING=TIME option (which is the default) displays additional information that includes ceiling times at the stages, and the CEILING=N option displays additional information that includes ceiling sample sizes at the stages.

SURVEYFREQ Procedure

The SURVEYFREQ procedure now produces simple and weighted kappa coefficients.

The new COLORSTAT= option for mosaic plots colors the tiles according to the values of the Pearson residuals. The CLDISPLAY=SERIFARROW and CLDISPLAY=LINEARROW options are now available to control the error bars in the odds ratio, relative risk, risk difference, and kappa plots.

The CELLCHI2, DEVIATION, and PEARSONRES options display the cell chi-square contributions, deviations, and Pearson residuals in the crosstabulation table.

SURVEYMEANS Procedure

If you request domain analysis by specifying a DOMAIN statement and you request estimates for percentiles or quantiles, the SURVEYMEANS procedure produces domain quantiles estimates. If you specify a POSTSTRATA statement, PROC SURVEYMEANS also incorporate s the poststratification in the domain analysis for quantiles.

You can request plots by specifying the PLOTS= option in the PROC SURVEYMEANS statement for continuous analytical variables. PROC SURVEYMEANS provides a summary plot that includes a box plot and a histogram plot. When you specify a DOMAIN statement, PROC SURVEYMEANS also produces box plots for domain statistics.

SURVEYPHREG Procedure

The DETAILS suboption of the VARMETHOD=BRR and VARMETHOD=JACKKNIFE options in the PROC SURVEYPHREG statement displays the maximum likelihood estimates of model parameters for replicate samples when the replicate parameter estimates are available. A replicate sample might not provide useful parameter estimates (replicate estimates), for reasons such as nonconvergence of the optimization or inestimability of some parameters in that replicate sample.

SURVEYREG Procedure

When the regression model depends on at most one continuous variable as a regressor, excluding the intercept, PROC SURVEYREG produces a fit plot.

SURVEYSELECT Procedure

The CERTSIZE option for specifying the certainty size measure is now available for METHOD=PPS_SYS and METHOD=PPS_WR. You can now specify the interval and start values for METHOD=SYS and METHOD=PPS_SYS.

The GROUPS= option requests random assignment of the observations in the input data set to groups. You can specify the total number of groups as *n*, which must be a positive integer. Or you can provide a list of group size values, which are positive integers that specify the number of observations in the groups.

Enhancements to the High-Performance Procedures

Beginning with SAS/STAT 12.3, available with Base 9.4, the high-performance procedures that are available with SAS High-Performance Statistics software for distributed computing are also available with SAS/STAT software for use in single-machine mode. These procedures are documented in *SAS/STAT User's Guide: High-Performance Procedures*.

HPCANDISC Procedure

The new HPCANDISC procedure performs high-performance canonical discriminant analysis.

HPFMM Procedure

The new HPFMM procedure performs high-performance finite mixture model analysis.

HPLMIXED Procedure

The OUTPUT statement creates a data set that contains predicted values and residual diagnostics, which are computed after the model is fit. The variables in the input data set are not included in the output data set in order to avoid data duplication for large data sets; however, variables that are specified in the ID statement are included.

The RANKS option in the PROC HPLMIXED statement displays the rank of the design matrix.

HPPRINCOMP Procedure

The new HPPRINCOMP procedure performs high-performance principal component analysis.

HPREG Procedure

The SCREEN option in the SELECTION statement requests screening stages that reduce a large number of regressors to a much smaller subset from which the final model is chosen. You have control over the displayed results, the number of stages, the number and percentage of effects to be chosen, and the cutoff value of the screening statistic.

What's Changed

The following sections describe changes in software behavior from SAS/STAT 12.1 to SAS/STAT 13.1.

FMM Procedure

PROC FMM now displays the effective sample size (ESS) table by default with Bayesian analyses. In addition, PROC FMM computes and displays the mixing probability for the kth component for simple probability models that are estimated with maximum likelihood.

FREQ Procedure

PROC FREQ now displays exact p-values, mid p-values, and point probabilities in the PVALUE6.4 format by default. You can control the format by specifying the new PFORMAT= option in the EXACT statement.

LOGISTIC Procedure

Specifying the ODDSRATIO statement no longer suppresses the default odds ratio table; to suppress the default table you can specify the NOODDSRATIO option in the MODEL statement. Specifying the CLODDS= option still suppresses the default odds ratio table.

MCMC Procedure

PROC MCMC has reduced the number of posterior statistics and convergence diagnostics that are displayed by default. The "Summary Statistics" and "Interval Statistics" are combined into a joint table. In diagnostics output, PROC MCMC now displays only the "Effective Sample Sizes" table. The "Monte Carlo Standard Errors," the "Autocorrelation," and the "Geweke Diagnostics" tables are generated by using the DIAGNOSTICS= option.

NLMIXED Procedure

PROC NLMIXED now constructs the random-effects clusters by using each of the unique SUBJECT= variable values. You can change the clustering behavior to that of previous releases by specifying the NOSORTSUB option in the PROC NLMIXED statement.

NPAR1WAY Procedure

PROC NPAR1WAY now displays exact p-values, mid p-values, and point probabilities in the PVALUE6.4 format by default. You can control the format by specifying the new PFORMAT= option in the EXACT statement.

References

Fine, J. P. and Gray, R. J. (1999), "A Proportional Hazards Model for the Subdistribution of a Competing Risk," *Journal of the American Statistical Association*, 94, 496–509.

Schemper, M. and Henderson, R. (2000), "Predictive Accuracy and Explained Variation in Cox Regression," *Biometrics*, 56, 249–255.

Train, K. E. (2009), *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.